More than meets the eye: an eye-tracking study of the effects of translation on the processing and memorisation of reversed subtitles

Valentina Ragni, University of Leeds

ABSTRACT

Reversed subtitling is a subtitling mode in which the original audiovisual product is in the native language of the viewer, and the subtitles are in the foreign language (L1 audio, L2 subtitles). Distinct acquisitional advantages of reversed subtitles have emerged from previous research, especially for L2 vocabulary retention, a crucial component of foreign language learning. Despite these benefits, however, reversed subtitles remain one of the least explored subtitling modes to date, with a particularly acute lack of knowledge regarding how they are processed and their impact on memory. The aim of this paper is to resume and expand the discussion on this type of interlingual subtitles, and fill this gap by presenting the initial results of a study on the effects of reversed subtitle translation on reading and retention of L2 input. Specifically, formal similarity (literal transfer) and formal discrepancy (non-literal transfer) were compared. To determine how the two translation conditions were processed, eye tracking was used. To ascertain their effect on retention, an immediate L2 recognition post-test was administered. The participants were English (L1) native speakers learning Italian (L2) at an upper-intermediate level (CEFR B2). Findings show that the reversed subtitles were processed, a large percentage of L2 input was retained after a single exposure to the video, and translation-specific factors are liable to affect the learner.

KEYWORDS

Subtitling, reversed subtitles, audiovisual translation (AVT), eye tracking, cognition, subtitle processing, second language acquisition (SLA).

1. Introduction

The development of Information and Communication Technology (ICT) together with the pervasiveness of the internet have allowed information to flow far and wide on a global scale. A large portion of this information is audiovisual in nature, especially since the advent of smart phones and the introduction of faster and cheaper data plans for mobile communications. The idea of using the subtitles that often accompany this dazzling array of audiovisual products as aids in the teaching and learning of languages is not new, having been investigated academically since the 1980s. The topic has been gaining steady momentum particularly over the past 10-15 years, during which it has become established that the addition of pre-recorded subtitles on video has positive effects on various aspects of learning a foreign language (FL or L2). Subtitles can be classified as intralingual (also called monolingual in Translation Studies, and bimodal input in Psychology) or interlingual (i.e. involving translation from one language into another). Interlingual subtitles can be further classified into standard and reversed, depending on where the FL appears. Standard subtitling can be defined as a subtitling mode where a FL audiovisual product is translated into the

native language of the viewer (L2 audio, L1 subtitles). In reversed subtitling, on the other hand, the original audiovisual product is in the native language of the viewer, and the subtitles are in the foreign language (L1 audio, L2 subtitles).

Compared to standard (L2 audio, L1 subtitles) and bimodal input (usually L2 audio, L2 subtitles), reversed subtitling still remains relatively underexplored, despite several studies (reviewed in detail in section 2) having highlighted its effectiveness in assisting the learner in the processing and learning of L2 input presented audiovisually. After reviewing some of the reasons for this effectiveness, Díaz-Cintas and Fernández Cruz (2008: 215) rightly pose the following question: "why are reversed subtitling and bimodal input not very common teaching techniques?". Ten years later, the situation remains unchanged: very few studies focus on this subtitling mode, and its use in the foreign language classroom is still sparse.

In order to see a tangible increase in the use of subtitled video in foreign language learning and teaching (FLL and FLT, respectively), an array of learning activities exploiting this audiovisual mode need to be created and disseminated to teachers. This paper argues that, to be able to develop appropriate learning activities using reversed subtitling, more experimental research is first needed to establish the underlying cognitive mechanisms at play during the processing of this particular type of L2 written input, and also its specific acquisitional benefits. The study presented here takes a step in this direction by addressing reversed subtitles directly (rather than comparatively, like most research available to date) and considering an aspect of L2 input that, in the field of second language acquisition (SLA), is rarely acknowledged to have an effect on the learner: translation. This paper expands the limited knowledge of reversed subtitling by assessing translational factors in the processing and memorisation of L2 input during the consumption of filmic products, thus adding to the growing body of psycholinguistic literature on audiovisual translation (AVT). Specifically, formal similarity and formal discrepancy created via translation are the focus of this investigation, and eye tracking is used to shed light on patterns of attention allocation to these translation conditions.

First, the handful of studies dealing with reversed subtitling are reviewed. Starting from psycholinguistics and research on reading using eye tracking, factors that play a role in the perception of AVT are addressed. The concept of formal similarity is explained and evidence of its impact on the learner is drawn from both SLA and AVT, before introducing the experiment. The translation protocol provides detailed explanations of how the experimental stimuli were devised. The methodology describes participants, study design, instruments, research questions and experimental procedure. The data analysis answers two core research questions, and the discussion brings the findings together and analyses their implications. The last sections addresses some limitations of the study, suggests avenues for further research and draws some conclusions based on the findings.

2. Research on reversed subtitles

Seminal work starting in the 1980s demonstrated that reversed subtitles were beneficial for memory retention (of both exact phrasing and contextual meaning), comprehension, L2 encoding, spelling, vocabulary learning, and activation of prior knowledge (Lambert *et al.* 1981; Lambert and Holobow 1984; Holobow *et al.* 1984; Danan 1992). Except for Danan, however, in these studies 'audiovisual' meant involving two verbal modes only, audio and subtitles. Different combinations of spoken dialogue and written script (subtitles) were manipulated, but no moving images (visual nonverbal mode) were included, so the studies did not test actual video.

In the 1990s, the Belgian school addressed reversed subtitles in some of their studies comparing several combinations of audio and text, and found solid acquisitional advantages of reversed over standard subtitles for vocabulary learning (Pavakanum and d'Ydewalle 1992; d'Ydewalle and Pavakanum 1996, 1997; d'Ydewalle and Van de Poel 1999; Van de Poel and d'Ydewalle 2001). These studies included three presentation modes (audio, subtitles, images), but the experimental materials were 'still-motion movies', i.e. successive static pictures and not film as we know it today. The moving images change both the viewing experience as a whole, which becomes more dynamic (a continuous flow rather than a succession of separate shots), and the subtitles (which need to work around fast shot and scene changes in this seamless video stream to avoid flickering), therefore potentially also changing attention distribution patterns and subsequent retention of L2 features. SLA and psycholinguistic studies show that "audio-visual integration is an inherent part of bilingual language processing, be it integration of auditory and visual *modalities* or integration of phonological and orthographic *word-forms*" (Marian 2009: 70-71, my emphasis). Therefore, linguistic factors are not the sole source of effects on the viewer/learner, but the presentation modes themselves are liable to affect the mapping of L2 input in reversed subtitles to its L1 aural counterpart.

To the author's knowledge, the only studies where reversed subtitling is investigated in 'true video', i.e. moving images, are the ones conducted by Baltova (1999), Van Lommel *et al*. (2006), d'Ydewalle and De Bruycker (2007), Zarei (2009), Čepon (2011, 2013), and Bisson *et al*. (2014), but evidence is inconclusive as the results do not yet construct a full, coherent picture of processing and acquisition in this mode. Perhaps the two most comprehensive studies on reversed subtitles are the last two, which exploited eye tracking in their investigations, a technique used for at least four decades in reading research to analyse the processing of written text (Keating 2014). In their study comparing standard and reversed subtitles for Dutch L1 speakers with no knowledge of Swedish L2, d'Ydewalle and De Bruycker (2007) found a higher skipping rate for reversed subtitles, more 'normal' reading patterns with standard subtitles (i.e. they better approximated reading static text), a less consistent attention distribution

with reversed subtitles (25% of the watching time spent fixating reversed subtitles, 35% standard subtitles), longer latencies in the reversed condition, especially for two-liners, and lower mean fixation duration in reversed compared to standard subtitles (significant for two-liners). Bisson *et al.* (2014) compared native English (L1) speakers with no knowledge of Dutch (L2) on four conditions: L2 audio only, standard, bimodal, and reversed. With reversed subtitles, they also found significantly higher skipping rates, and significantly lower fixation counts and durations. Despite subjects not *needing* to process the reversed L2 subtitles (owing to the presence of L1 soundtrack), they nevertheless spent a considerable amount of time reading them, as indicated by the average normalised fixation count (0.59), meaning they fixated on more than half the words in the reversed condition, even if the L2 was completely unknown to them (*ibid*.: 413). Bisson and colleagues included an acquisitional measure (vocabulary test), and despite the lack of main effects of subtitle condition on vocabulary recall, t-tests showed that the mean recall scores for all conditions were significantly higher than chance, suggesting some retention of vocabulary after a single exposure to reversed subtitles did occur, despite the lack of familiarity with the L2. The more erratic eye movement patterns reported for reversed subtitling are unsurprising because in both studies the L2 (Swedish and Dutch, respectively) was completely meaningless to the viewers. The situation is bound to be different with L2 learners, who can be assumed to be interested in the L2 and therefore have more reasons to look at the subtitles. Consistent with earlier research, this paper addresses processing of reversed subtitles in true learners of the foreign language in question, Italian (see 6.1).

3. Lessons from psycholinguistic research

The two studies reviewed above are useful because they are the first empirical explorations of reversed subtitle processing, and they allow researchers to start building a picture of what AVT processing in this viewing condition looks like. As they show, eye tracking can add a powerful dimension to FLL investigations because it taps into less conscious processes such as early stages of word recognition (Van Assche *et al.* 2009) and the integration of new information into the current sentence during reading (Cop *et al.* 2015).

However, if what is of interest, as in the present study, is the usefulness of reversed subtitles *from an acquisitional perspective*, a measure of language learning should be included alongside processing. The only study that has done this for reversed subtitling so far, to the author's knowledge, is Bisson *et al.* (2014). Despite d'Ydewalle and De Bruycker's (2007) being one of the most thorough investigations of reversed subtitles to date, the study failed to clarify the relationship between reading and attentional patterns and performance measures of language acquisition (Kruger 2016: 7). In the authors' words:

> An experiment that directly links a detailed analysis of eye movements in reversed subtitles with foreign language acquisition is needed in order to make more conclusive inferences on the linguistic processing of foreign language subtitles in the presence of a native language soundtrack. (d'Ydewalle and De Bruycker 2007: 204)

The current study bridges this gap precisely by considering behavioural and acquisitional measures together, exploring the relationship between physiological characteristics of attention allocation (eye movements) and mnemonic performance (recognition and recall) on L2 items for English learners of Italian. Fixations specifically are analysed, as they are good indices of initial lexical processing (Rayner and Liversedge 2011), where longer fixations correspond to higher cognitive effort, as readers pause on words that require more processing (Just and Carpenter 1980).

A host of factors influence eye movements during reading, such as word frequency (Inhoff and Rayner 1986; Rayner and Duffy 1986), predictability and contextual constraints (Balota *et al*. 1985; Kliegl *et al.* 2004), familiarity (Williams and Morris 2004), polysemy and lexical ambiguity (Sereno *et al*. 2006), morphology (Hyönä and Pollatsek 1998), including the morphological decomposition of compound words (Pollatsek and Hyönä 2005), but also syntactic complexity and other semantic and pragmatic factors (Clifton *et al*. 2007; Morris 1994), as well as reading skills (Rayner 1986) and age of acquisition (Juhasz and Rayner 2003, 2006).

In AVT, a number of linguistic and paralinguistic factors have been identified as liable to influence processing alongside the presence of video and audio, namely text reduction techniques such as key-word captioning (Guillory 1998; Montero Perez *et al.* 2014), word omission (De Linde and Kay 1998), subtitle segmentation and text chunking (Perego *et al*. 2010; Rajendran *et al.* 2013; Kruger *et al.* 2015; Gerber-Morón and Szarkowska 2018), number of lines (D'Ydewalle and De Bruycker 2007; Kruger and Steyn 2014), video genre (D'Ydewalle and Gielen 1992), word frequency and cohesion (Moran 2008, 2012). Despite being in its infancy, psycholinguistic research in AVT is experiencing a dramatic surge in popularity, which makes it all the more surprising that translation itself, such an intrinsic and visible aspect of interlingual subtitling, has virtually never been tackled directly. The next section addresses some of the reasons why translation, the strategies used to produce it, and issues of L1-L2 equivalence may play a role in the consumption of audiovisual material, making a case for why this topic deserves more attention in reception studies in AVT, both theoretically and experimentally.

4. Formal similarity in translation

In this paper, formal similarity between L1 and L2 in reversed subtitle translation is under scrutiny. Formal similarity is intended broadly as a situation where "two languages use the same linguistic elements to express the intended meaning, while formal difference occurs when the meaning is

expressed by different linguistic means" (Laufer 2000: 187). Crucially, therefore, the term 'formal' is not to be intended purely as surface form overlap but includes instances of broader syntactic L1-L2 correspondence, as well as established translation equivalents. The latter are cases in which an L2 word or expression is the standard, most commonly accepted translation of its L1 counterpart, i.e. there is semantic and functional equivalence even if the two items do not share surface form. Examples of formal similarity between Italian and English are the lexical pairs *ristorante*–*restaurant* (where there is some degree of surface form overlap), *amico*–*friend* (no surface form overlap but semantic and functional equivalence), and the syntactic pairs *Hai mangiato?* [Have you eaten?]–*Have you eaten?* and *Quanti anni hai?* [How many years do you have?]–*How old are you?* (for full details and more examples, see 5).

Investigating formal similarity and its flipside, discrepancy, in translation is of interest for several reasons. First of all, subtitle translation may have an impact on processing and learning as it is responsible for the activation of comparison processes between L1 and L2 (Ghia 2012b), and this may be particularly true in advanced FL learners, who may be more likely to make effective moment-by-moment comparisons, as their higher proficiency allows them to process and comprehend the FL input more quickly, freeing cognitive resources that can be dedicated to particularly salient L2 items. Translation is also relied on when FL skills are less developed, as new L2 input is more strongly linked to its L1 equivalent at lower proficiency levels (Prince 1996). Translation is increasingly being recognised as a spontaneously occurring cognitive activity (Pavesi 2002; Leonardi 2010; Caimi 2005, 2012) and it appears to be not just naturally occurring, but also naturally interesting. Evidence for this comes from learner comments found in Čepon's (2011: 15) study on English L2 acquisition through reversed subtitles, where a student stated: "reading subtitles in English was quite interesting because I was curious how our words were translated so I was constantly comparing Slovenian and English; and that was certainly more fun than I expected". In Ragni (2017), several similar comments showed that university language students make explicit reference to their translation skills and use the subtitles to check for translation solutions, even if they are not studying translation specifically. It is therefore reasonable to assume they will pay some attention to reversed subtitles, potentially noticing discrepancies or similarities in the L2 input during processing, which could in turn lead to retention.

Secondly, there is ample evidence from cross-linguistic influence (CLI) research that language similarity affects both L2 processing and learning. For instance, L1 orthographic backgrounds play a role in L2 visual word recognition and lexical processing (Wang *et al.* 2003; Koda 1996, 1997, 1999; Muljani *et al.* 1998) as well as L2 word-form learning (Laufer 1997; Hamada and Koda 2011). Cognate studies provide evidence that bilinguals read cognate words faster than control words because of shared features at the word and sub-word level (Dijkstra *et al.* 1999). In the initial stages,

learners rely on lexical associations from L2 to L1 and therefore can be prone to confusion when encountering orthographically similar words, especially if they do not share semantics as in the case of false friends. This situation progressively changes as proficiency increases, with learners becoming more likely to mediate L2 words through concepts, although mediations at the lexical level do not disappear altogether but remain as a form of interlanguage connection (Talamas *et al*. 1999). Significant L1-L2 overlaps such as cognates also affect the development and organisation of the L2 mental lexicon in early stages of vocabulary development (Hall 2002). Moreover, whether they are cognates or not, translation equivalents, such as *photography-fotografia* and *book-libro* in English-Italian, "appear to have a different and closer cognitive status than within-language synonyms" (Francis 2005: 251), perhaps because these immediate lexical associations at the representation level are established upon the first encounter with a new L2 word. Thus, L2 learners economically maximise already established language structures (i.e. entries in the mental lexicon) to connect L2 words to a translation equivalent, and this may be particularly evident in a situation like that of interlingual subtitling, where L1 and L2 are available simultaneously and throughout.

In Translation Studies, since the early age of pre-linguistics writings on translation, opinion has "swung between literal and free, faithful and beautiful, exact and natural translation" (Newmark 1981: 38). In fact, the distinction between word-for-word and sense-for-sense (i.e. literal and free) dates back to Cicero and St. Jerome (Munday 2008), but remains unexplored experimentally. In AVT more specifically, CLI and the effects of formal similarity have virtually never been tackled directly. However, scholars have occasionally observed them in the discussion of their findings, both at the vocabulary level (Guichon and McLornan 2008) and broader language family level (d'Ydewalle and Pavakanum 1997; Van de Poel and d'Ydewalle 2001).

The only notable exception to this trend so far is a study by Ghia (2012a), who specifically addressed translation effects created via input enhancement on learners' L2 noticing and memorisation in standard subtitling (English L2 audio, Italian L1 subtitles). The term *input enhancement* refers to how a translational contrast is created by highlighting formal divergence of input through translation strategies such as reduction and substitution. To investigate *translational salience*, namely "the prominence acquired by linguistic items when input is delivered" (*ibid*.: 52), she compared literal (formally similar) and non-literal (formally diverging) L1 subtitles and used eye tracking. Literal translations involved no main L1-L2 changes, with word order and vocabulary preserved between audio and subtitles, often resulting in word-for-word renderings, while diverging, non-literal translations involved lexical and syntactic substitutions, pragmatic substitutions and reductions (*ibid*.: 78). The memory post-test consisted of a multiple-choice questionnaire (MCQ) of verbatim recognition of L2 wordings as heard in the foreign audio. As for

eye movements, Ghia analysed mainly shifts between images and the subtitles and regressions within the subtitles. Although no difference was found for regressions, shifts were significantly higher with discrepancy, and second-pass fixations were directed to diverging content words. This intensified processing occurring in non-literal subtitles suggests that some form of discrepancy detection was in place, whether conscious or not. Cognitively speaking, it seems thus reasonable to postulate that there is a difference in source-target mapping in the two conditions. According to Tanenhaus (2007), increased mapping generally indicates a higher degree of attention to the verbal dimension of input, i.e. to the words being heard or read, as opposed to other elements of the video such as proxemics or gesture. One might therefore assume that mapping is intensified in the diverging condition, where visual attention and processing are higher. However, this difference did not lead to improved memory performance: overall recall scores were higher in the literal condition, both for vocabulary and syntax. Interestingly, Ghia also found that, in most cases of misrecognition of non-literal items, learners chose the literal back-translation of the subtitles, suggesting that L1 interference occurred, resulting in higher focus on L1 input to the detriment of L2 mnemonic accuracy, which is consistent with findings from other studies on standard subtitles (Guichon and McLornan 2008; Mitterer and McQueen 2009). If, as we have seen, translation equivalence in psycholinguistic terms has a closer cognitive status than simple synonyms in the learner's mind, and L2 learners use existing entries in the mental lexicon to connect L2 words to a translation equivalent via lexical links, it makes sense that when translations deviate "from what could have been perceived as the 'norm' (i.e. literal translation)" (Ghia 2012a: 82), this correspondence between translation pairs is violated and the cognitive perception of this mismatch results in inhibitory effects on recall.

The above review suggests that formal similarity (or lack thereof) may impact on the processing and subsequent retention of L2 input during reversed subtitle reading too. The experiment presented hereafter, modelled on Ghia's, is conceived as a proof-of-concept study to assess whether the similarity vs. discrepancy distinction actually has a psychological reality in terms of perception and memory, and whether it can therefore be a viable route to investigate the effect of linguistic differences on the viewer. Before addressing the study methodology, the translation protocol details how formal similarity and discrepancy were created in the subtitles.

5. Translation protocol

Formal similarity is closely linked to the concept of 'literal transfer' as postulated by Gottlieb (1992), i.e. "a translation that reproduces the original text as much as possible both in lexical and syntactic terms, and occasionally involves word-for-word rendering" (Ghia 2012b: 167). Therefore, while formal similarity refers to one of the two experimental

conditions in this study, literal transfer refers to the specific translation strategy used to produce the formal similarity condition.

Owing to typological differences between Italian and English, a word-for-word translation, in particular of syntactic structures, is not always possible. Literal transfer therefore does not mean total identity of both form and meaning. Given that absolute L1-L2 lexical and syntactic identity does not often (if ever) exist, the term 'equivalence' has been traditionally used in Translation Studies. For the purpose of this paper, equivalence does not refer to any specific translation theory but is used as intended by Hervey *et al*. (2000: 19), namely "in its everyday sense of counterpart – something different, but with points of resemblance in relevant aspects." Here, therefore, *formal* similarity does not solely refer to *orthographic* equivalence or quasi-identity of individual words and word-for-word translation but includes instances of semantic and functional equivalence of individual words as well as structural correspondences between L1 and L2. Examples 1 and 2 below show instances of a word-for-word translation. In all the following examples and in the remainder of the paper, L stands for formal similarity (literal translation), whereas N stands for formal discrepancy (non-literal translation).

Example 1

ST: But an oracle can?	
L: Ma un oracolo può? [But an oracle can?]	N: E l'oracolo invece sì? [And the oracle instead yes?]

Example 2

ST: You have a good soul.	
L: Hai un buon animo. [You have a good soul.]	N: Hai un buon cuore. [You have a good heart.]

Example 2 shows a case of lexical intervention because only the L1 item 'soul' was changed between L and N, whereas Example 1 shows a case of syntactic change because the whole structure of the subtitle was addressed in the creation of the two translation versions. The extent of the change has been underlined in all examples.

In Example 3, on the other hand, Italian simply does not have the same construction with the adverb 'right'. However, 'to be on time' translates literally as *essere in orario* [to be in time], which is the most commonly accepted translation. The qualifier moved from adverbial into adjectival position and was rendered with a collocate of *in orario*. This syntactic literal translation is fully functionally equivalent and much closer to the original in terms of syntactic structure than other alternatives like *sei davvero puntuale* [you are really punctual] or *sei arrivato all'ora stabilita* [you have arrived at the agreed time], and therefore constitutes an instance of formal similarity.

Example 3

ST: You're right on time.	
L: Sei in perfetto orario. [You are in perfect time]	N: Sei davvero puntuale. [you are really punctual]

Once a literal translation was identified, formal discrepancy was achieved through a second, freer translation. Different strategies were used to create formal discrepancy depending on whether lexicon or syntax was concerned. An instance of lexical divergence occurs in Example 4, where the literal translation of 'prophecy', *profezia*, was substituted with the term 'vaticination', which has the same meaning and fits the context but differs radically in surface form ('prophecy' and *vaticinio* do not share orthography):

Example 4

ST: …the prophecy?	
L: la profezia? [the prophecy]	N: il vaticinio? [the vaticination]

Syntactic divergence can be observed in Example 5, where *fare da guida* [act/serve as a guide] instead of the literal *essere una guida* [to be a guide] introduces a change in sentence structure but allows the retention of the central focus of the utterance; the noun *guida*, in both versions, maintains a similar length and rhythm, and has the same number of words as L:

Example 5

ST: She is a guide, Neo.	
L: Lei è una guida, Neo. [She is a guide, Neo.]	N: Ti farà da guida, Neo. [She will be your guide, Neo.]

6. Methodology

6.1 Participants

Participants were native English speakers aged 20-35 (M = 23.8; SD = 3.9) learning IFL at an upper-intermediate level (CEFR B2) with normal or corrected-to-normal vision. The gender ratio was 18:8 (69% female, 31% male), which reflects the bias towards female gender typical of foreign language learning demographics. Tweny-six people took part in the study (13 per group). Participants were BA students of Italian at the University of Leeds, UK, in their last or second-to-last year, MA students, or young professionals who had graduated less than five years before the experiment and used Italian on a weekly basis. All participants had been studying Italian for at least four years and passed the IFL test (CEFR B2 level). The relative variety of this demographic means the sample analysed is not limited to university students aged 20-23 but is more representative of young adults in general.

6.2 Design

The study adopts an experimental 2x2 Latin square design (Table 1), where two viewing conditions are compared: literal translation (L) and non-literal or diverging translation (N):

	Translation Condition	
	Part 1	**Part 2**
Group 1 (G1)	Literal (L)	Non-literal (N)
Group 2 (G2)	Non-literal (N)	Literal (L)

Table 1. Counterbalanced study design

Participants who were randomly assigned to G1 watched the first half (Part 1) of the video in L (literal condition, i.e. formal similarity) and the second half (Part 2) in N (non-literal condition, i.e. formal discrepancy). The reverse applied to G2. In this design, each subject's behaviour is explored under both conditions, both halves are subtitled in both conditions, and each stimulus acts as its own control (Godfroid *et al.* 2010).

The independent variables are translation condition (L or N), fixation duration and fixation count, whereas the dependent variable is performance in the verbatim recall post-test.

The experimental stimulus was a 10 min 40 sec excerpt from the film *The Matrix* (Andy Wachowski and Lana Wachowski, 1999). A total of 110 subtitles were analysed, 55 in each half. The subtitles were created in the professional software Softel Swift 6.0. Subtitle in-times and out-times remained the same between G1 and G2, ensuring that attention allocation was not affected by differences in subtitle duration on screen. The only element that changed between G1 and G2 was the Italian translation contained in the subtitles.

The two translation conditions differed by one, two or three words maximum, and often it was possible to maintain the same number of words between L and N. The clip was displayed with a black letterbox, where all subtitles were presented in a centre-aligned position, as illustrated in Figure 1[1]:

Figure 1. Example of letterbox on the experimental stimulus

Subtitles ranged between 2-70 characters, including spaces and punctuation. They were between 1-13 words long, either one- or two-liners, and chunked at grammatical and logical points in the sentences (Rajendran *et al.* 2013; Karamitroglou 1998; Díaz-Cintas and Remael 2007). Characters per line did not exceed 37, and maximum reading speed was 180 WPM (words per minute).

6.3 Instruments

An immediate subtitle recognition multiple-choice question (MCQ) post-test was designed to capture potential differences in short-term verbatim memory stemming from L1-L2 discrepancies. It is a verbatim test because it asked participants to recognise the exact phrasing of the Italian subtitles they saw during the video. The post-test addressed both lexicon and syntax (see 5) and consisted of 22 items (subtitles) in total, 11 appearing in Part 1 and 11 in Part 2. Test items were chosen based on the relevance and clarity of the literal vs. non-literal translation distinction. To limit the confounding effect of guessing, a distracter was presented alongside the L and N versions. As in Ghia (2012a), distracters were *de facto* non-literal renderings. Distracters were chosen on a case-by-case basis, depending on source-target culture considerations, context, level of overlapping of semantic spheres, availability of synonyms or near-synonyms in Italian, sentence structure, discourse, register, word frequency, as well as other, not strictly grammatical considerations, such as confounds derived from question layout. A complete example of post-test item is question 18 below (Q18), modelled on the subtitle presented in Example 6:

Example 6

ST: I know you’re Neo.	
L: So che sei Neo. [I know you're Neo.]	N: Tu devi essere Neo. [You must be Neo.]

Q18: When Neo is finally received by the Oracle, what’s the first thing she tells him as he enters the room?

A – Neo, giusto?	[Neo, right?]	[distracter]
B – So che sei Neo.	[I know you're Neo.]	[L]
C – Tu devi essere Neo.	[You must be Neo.]	[N]

An open questionnaire was also administered to measure free recall, metalinguistic awareness, and degree of noticing of the above-mentioned discrepancies created through translation manipulation. The questionnaire was administered online, immediately after the MCQ.

A standardised, written IFL (Italian as a Foreign Language) proficiency test was administered in paper form to control for confounds derived from individual differences in L2 competence. The exercises were taken from CEFR level B2 tests in the Certificazione di Italiano come Lingua Straniera, (CILS, i.e. Certification of Italian as a Foreign Language) and from the University of Verona’s Language Centre (Centro Linguistico di Ateneo). CILS is the only language qualification for foreign speakers of Italian recognised by the Italian government, and a requirement in Italian Universities. It is CEFR-standardised and part of the European Association for Language Testing and Assessment (EALTA). Exercises were extracted from official tests as these have been tested for validity and reliability through item analysis, inter- and intra-rater reliability (Barni *et al.* 2009). Not all areas covered in these official tests are relevant to the study (e.g. listening comprehension and speaking skills are not applicable given the AVT mode chosen and the nature of the study), therefore only relevant exercises testing reading comprehension and grammatical competence were extracted. Vocabulary, grammar structures and written text comprehension were assessed in the IFL test.

A standardised working memory (WM) test, namely the reading span task (RSPAN) first devised by Daneman and Carpenter (1980), was administered to control for confounds derived from individual differences in mnemonic ability. During this task, participants read aloud sets of sentences while simultaneously having to retain the final word of each sentence. At the end of each set, they try to recall these sentence-final words. Set size grows such that the number of final words subjects have to retain grows, increasing the processing effort required. The RSPAN adopted here used the same 42 test sentences used by Harrington and Sawyer (1992), with sets ranging between two and five sentences. Immediately after reading a sentence, participants took a grammaticality judgement task, where they had to say whether the sentence they had just read was a grammatical

statement. The RSPAN was administered as pen and paper, took around 30 minutes to complete and was timed to ensure self-paced reading out loud was comparable between participants. It was also taken right at the end of the experiment, to ensure participants would not know that their memory was being tested until they had already taken both the MCQ and the questionnaire.

Both by-subject and by-item variables were controlled for. The former include age, WM and IFL proficiency. The latter, subtitle length, duration, and frequency. Due to space restrictions, control variables are not addressed in this paper, but full details can be found in Ragni (2017).

6.4 Research Questions

Two research questions are addressed:

> **RQ1.** Which of the two translation conditions (L or N) yields a better recall rate in the verbatim memory post-test?
>
> **RQ2.** What can eye movements (fixation duration and count) tell us about the processing and memorisation of L2 subtitle translations?

RQ2 is divided into two sub-questions:

> > **RQ2a.** Do different attention allocation patterns emerge between the literal and non-literal condition?
> >
> > **RQ2b.** Is there a relationship between visual attention and recall rates?

Based on the findings of the only comparable study of translation effects by Ghia (2012a), discussed in section 4, we expected increased visual attention to translation discrepancies, i.e. more and longer fixations on non-literal subtitles, but formal similarity (literal translations) to be better recalled in the post-test.

6.5 Procedure

After signing a consent form, participants took the IFL test, then moved to the viewing station, where they sat around 60-70cm from the screen and, after a nine-point calibration, were instructed to watch the clip naturally whilst minimising body movements. Their eye movements were recorded with a Tobii TX120 eye-tracker running Tobii Studio software.

Participants were told that there would be general questions at the end about the content of the clip and their viewing experience as a whole, but did not know their memory would be tested. Immediately after watching, they sat the surprise verbatim memory test, followed by the questionnaire.

Then, they had a 5 minute break before sitting the WM test. All tests were self-paced but generally took around 20-30 minutes to complete.

7. Data analysis

A total of 33 participants were recruited, but seven subjects were excluded from the experiment because either they did not pass the IFL test; had an exceptionally high working memory;[2] possessed too high a working memory; had an eye condition; turned out to be bilingual from birth or because the eye-tracker did not satisfactorily record their eye movements. Eye-tracking data was analysed from 26 participants in total (13 per group), on all subtitles (n = 110) and on the test-items specifically (n = 22). All tests were run in the statistical software R (R Development Core team, 2017). By-subject and by-item analysis were carried out and included both descriptive and inferential statistics.

7.1 Recall and translation (RQ1)

Overall, 17 participants out of 26 (65.4%) recalled items more accurately when presented in L. Only four subjects recalled non-literal items (N) more accurately, and five obtained equal recall scores in both translation conditions. Descriptive analyses confirmed the data is normally distributed and there are no extreme outliers. The paired t-test shows a highly significant difference (95% CI: 0.74, 2.79; t = 3.5, df = 25, p = 0.001) between mean recall accuracy for literal (M = 8.84, SD = 1.64) and non-literal (M = 7.07, SD = 2.01) items. The t-test has a large effect size (Cohen's δ = .96) and power .88.

The above by-subject results were confirmed in the by-item analysis. The vast majority of subtitles (16 out of 22, i.e. 73%) were remembered more accurately when translated literally. Only four items were better recalled in N, and two subtitles had equal recall in L and N. In the literal condition, an overall 80% accuracy (correct recall n = 230) was registered across participants and items, against a 64% accuracy in the discrepancy condition (n = 184). The total responses elicited were 572 (26 subjects x 22 subtitles), 286 per translation condition (see Table 2). Of these, 158 (27.6%) were incorrectly and 414 (72.3%) correctly identified:

		Recall		Totals
		Correct (Yes)	**Incorrect (No)**	
Translation condition	**Literal**	230	56	286
	Non-literal	184	102	286
Totals		414	158	572

Table 2. Summary table of recall accuracy by translation condition

From the scatterplot in Figure 2, one can gather how the shape and distribution of data changes between the two conditions: non-literal (red) data are scattered, literal (black) data are concentrated at the top of the graph:

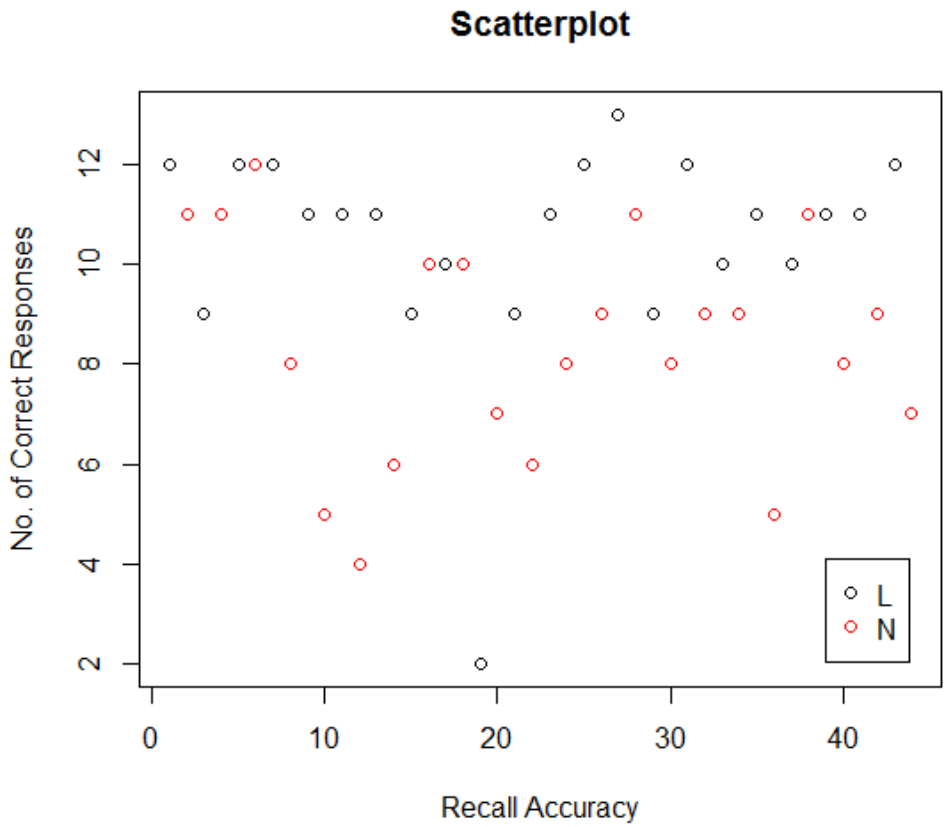

Figure 2. Scatterplot of recall accuracy by translation condition. All 44 renderings are included: the 22 literal renderings are marked in black, the 22 non-literal ones in red

The boxplots in Figure 3 visually reiterate that, overall, items were better recalled when translated literally. The top and bottom of the boxplot whiskers show that the full range of values is greater with diverging items (N). Thus, recall scores clearly varied more in the non-literal condition. The medians are in the middle of the box in both cases, suggesting the data is not skewed:

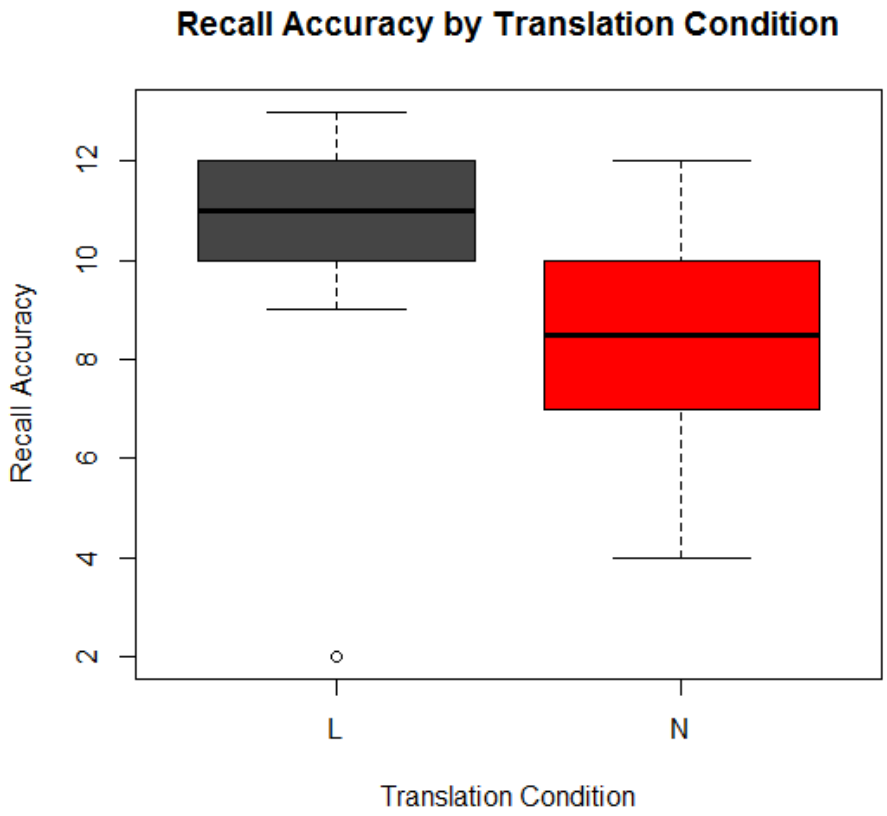

Figure 3. Boxplots of recall accuracy by translation condition. Literal (L) recall scores are marked in black, non-literal ones (N) in red

Descriptive analyses confirmed that data in L was slightly leptokurtic, but data in N was normally distributed, so the standard parametric t-test was used, as there was not enough evidence of non-normality. Mean differences were compared via the independent t-test, which shows a highly significant difference (95% CI: 0.73, 3.44; t = 3.1, df = 42, p = 0.003) between recall accuracy for literal (M = 10.45, SD = 2.21) and non-literal (M = 8.36, SD = 2.23) items, with a large effect size (δ = .93) and power .79.

7.2 Eye movements and translation (RQ2a)

Areas of Interest (AOIs) were drawn in Tobii Studio around the 110 subtitles in each of the two experimental clips (G1 and G2) and copied across to all participant recordings in that group, so that eye movements for all subjects assigned to G1 were recorded on exactly the same AOIs, and the same applied to G2. The total number of fixations on all 110 subtitles by all participants was 14,513. Items (i.e. subtitles) translated literally were fixated 7,176 times, non-literally translated ones 7,337. Each subtitle was fixated 65.23 times across participants in the literal and 66.7 in the non-literal condition. Descriptive analyses showed that fixation count data were not normally distributed, so the most common non-parametric alternative to the t-test was used, i.e. the Wilcoxon rank-sum test, also known as the Mann-Whitney test (Whitley and Ball 2002). On these data, the Mann-Whitney test showed no significant difference (95% CI: -9, 7; W = 5,953,

p > 0.5) between total fixation count on literal (*M* = 65.23, *SD* = 36.12) and non-literal (*M* = 66.7, *SD* = 35.83) items.

The relationship between translation condition and fixation count was then investigated, considering only the subset of 22 subtitles that constituted the post-test items. Total fixation count on these by all participants was 2,765. Items translated literally were fixated 1,307 times in total, non-literally translated ones 1,458. Fixation count on this subset of items was found to be normally distributed in all graphical and formal normality tests. As shown in Figure 4, the medians in L and N are close, so are the ranges and the maximum and minimum values. Post-test items were allocated a similar number of fixations in both conditions, and this difference is not significant (95% CI: -22.42, 8.69; t = -0.89, df = 42, *p* > 0.05).

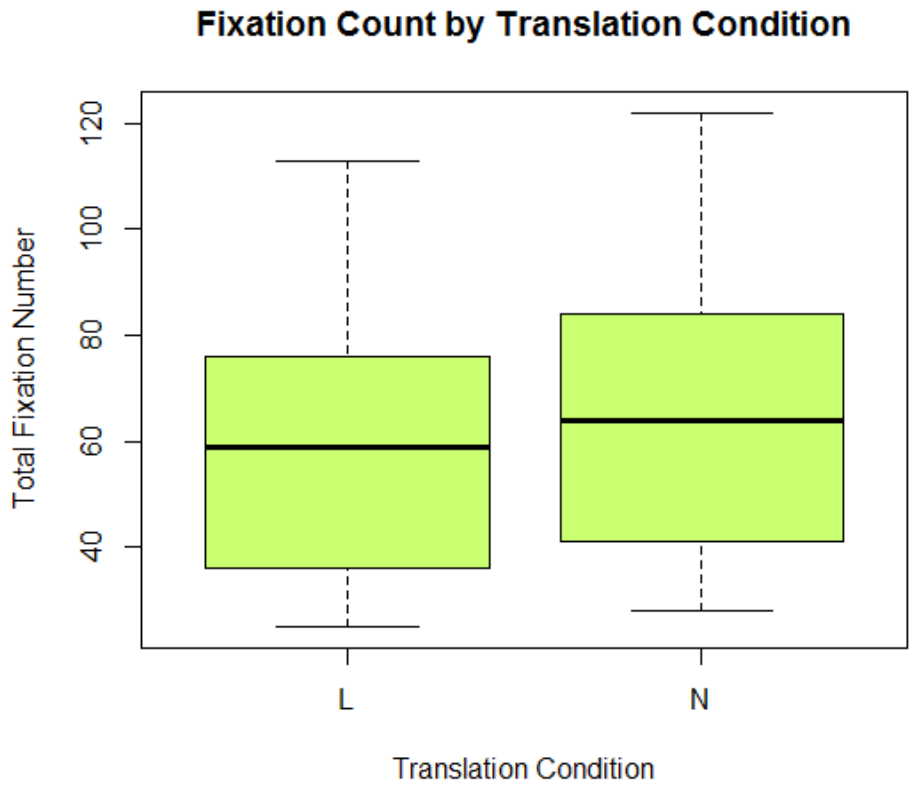

Figure 4. Boxplots of total fixation count on all subtitles by translation condition (22 subtitles)

As for fixation duration, overall mean fixation duration (MFD) on all 110 subtitles was 217 ms. MFD in L was 216 ms, in N 219 ms. A total of 26 min (1,552,511 ms) were spent fixating literal subtitles, against 27 min (1,607,414 ms) on non-literal ones. Therefore, items in N were fixated for longer than those in L, both in total and on average, yet the difference was not very large. Descriptive tests confirmed a departure from normality for fixation duration on both literal and non-literal items. The Mann-Whitney test showed no significant difference (95% CI: -1,948, 1,509; W = 5,939, *p* > 0.5) between total fixation duration on literal (*M* = 14,113 ms, *SD* = 7,365) and non-literal (*M* = 14,612 ms, *SD* = 7,681) items.

The relationship between translation condition and fixation duration was then investigated for the 22 post-test items. As shown in Figure 5, post-test subtitles were fixated for similar amounts of time in both conditions, and this difference is not significant (95% CI: -4,762.93, 1,771.47; t = -0.92, df = 40, $p > 0.05$).

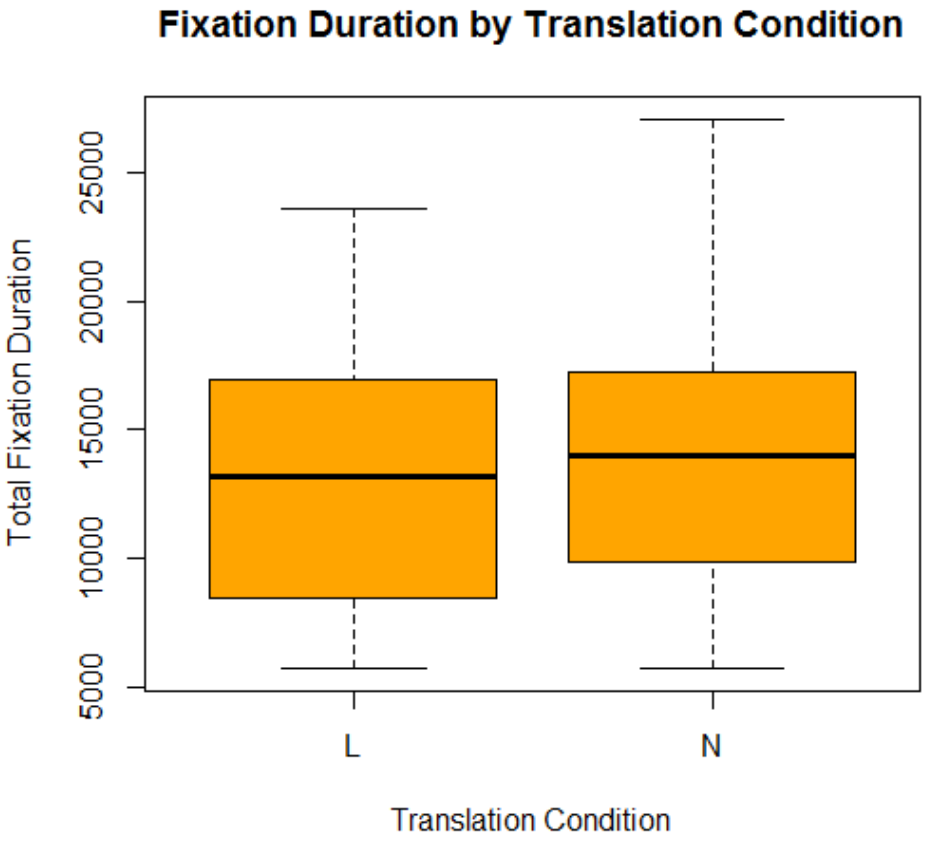

Figure 5. Boxplots of fixation duration on all subtitles by translation condition (22 subtitles)

A similar situation to that of fixation count was therefore found for fixation duration: the total number and length of the fixations made by participants did not change significantly depending on translation condition, either in the test subset (22 subtitles) or throughout the whole experimental stimulus (110 subtitles).

7.3 Eye movements and Recall (RQ2b)

As mentioned in 7.2, the 22 test items were fixated 2,765 times in total across participants. Of this total, correctly recalled items received 2,127 fixations, incorrectly recalled items only 638. Therefore, the more a subtitle is looked at, the higher the chances of it being remembered accurately. To get a more visual idea of the size of this difference in fixation count, the data are displayed in Figure 6, which visually confirms that correctly recalled items (Y) have a much wider range of values and received a much higher number of fixations compared to their incorrectly recalled counterparts (N):

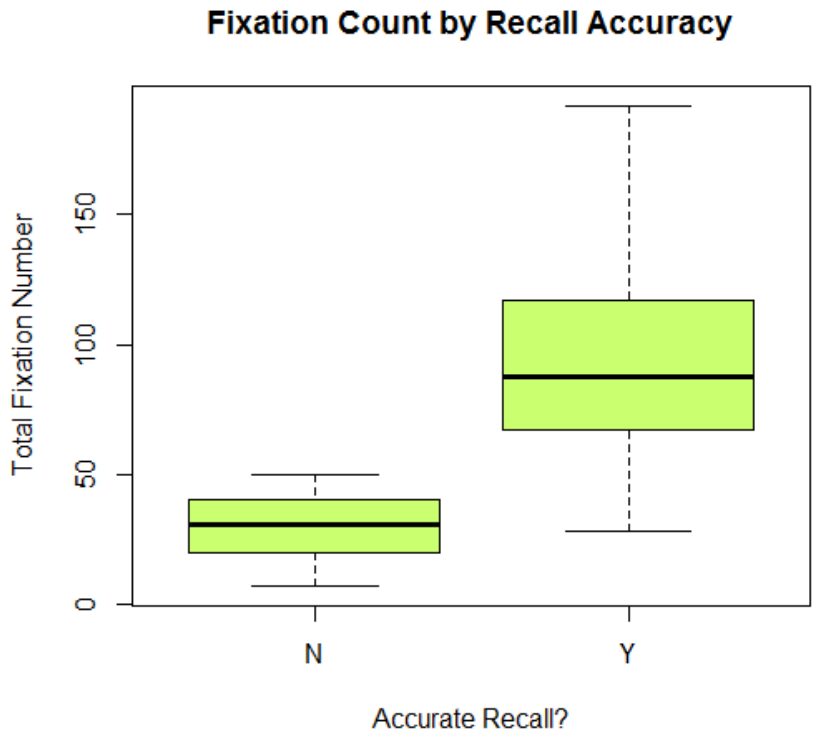

Figure 6. Boxplots of recall accuracy and fixation count. Y stands for 'yes' (correctly recalled items). N stands for 'no' (incorrectly recalled items)

A parallel situation arose for fixation duration, as illustrated in Figure 7:

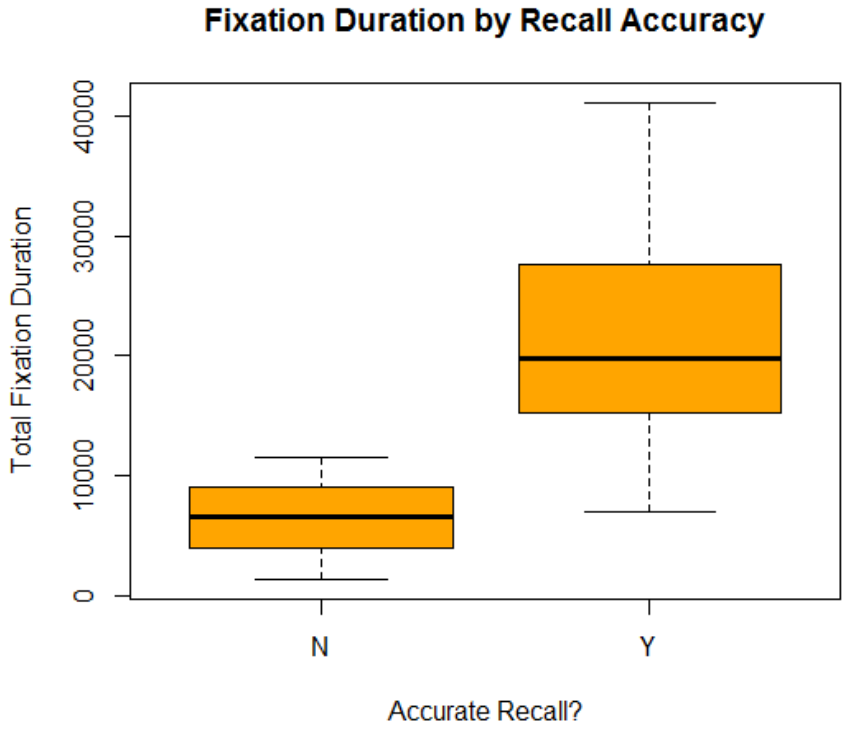

Figure 7. Boxplots of recall accuracy and fixation duration. Y stands for 'yes' (correctly recalled items). N stands for 'no' (incorrectly recalled items)

Fixation count data were normally distributed, but the Levene's test revealed heteroscedasticity ($F = 14.777$, $p < 0.0005$), so homogeneity of variance could not be assumed and the independent t-test default option with degrees of freedom adjustment (Welch correction) in R was therefore chosen. The independent t-test showed a significant difference in fixation count (95% CI: 46.75, 88.61; t= 6.66, df = 24.5, $p < 0.001$) between correct ($M = 96.68$; $SD = 45.72$) and incorrect ($M = 29$; $SD = 13.29$) items, with large effect size ($\delta = 2.0$) and power .99.

Fixation duration data was also normal but heteroscedastic as per Levene's test results ($F = 14.49$, $p < 0.0005$), so the Welch correction was applied. Results (95% CI: -19,145, -10,445; t = 6.99, df = 25.4, $p < 0.01$) confirmed that there was a significant difference in fixation duration between subtitles that were correctly recalled (M = 21,174.36, SD = 9,419.65) and those that were not (M = 6,379.27, SD = 3,099.35), with effect size 2.1 and power .90. The results mirrored those found for fixation count, suggesting that if items are fixated more times and for longer, they have a significantly higher chance of being remembered in the post-test.

8. Discussion

This study investigated the impact of formal similarity (L) and formal discrepancy (N) in translation on the processing and memorisation of reversed subtitles. The investigation was guided by two main research questions. The first (RQ1) asked which of the two translation conditions (L or N) led to a better recall rate in the recognition post-test. Statistical analyses revealed that, after one single exposure to reversely subtitled video, literal translations are recognised significantly more precisely than non-literal ones. It appears that formal similarity has indeed a psychological impact and can influence short-term recognition memory for exact phrasing.

A more in-depth analysis of the test-items revealed that certain subtitles were not only recalled correctly by the majority of the participants in L, but also had the poorest recall in N. The non-literal version of these items seems to create confusion, resulting in most people not being able to correctly recognise the discrepancy in the post-test, further reinforcing the superiority of the literal condition. Other items, however, especially when lexical in nature, were recalled quite well in N, showing that discrepancy can be mnemonically effective. Moreover, translation discrepancy was commented on extensively in the questionnaire. In fact, when asked to list any item from the clip that they found striking (free recall), many L and N items were mentioned but, perhaps somewhat counterintuitively, the majority were non-literal. This finding points towards a difference between free-recall and recognition, which may be based on interrelated but distinct processes (Clariana and Lee 2000), where recall involves further elaboration of memory traces, while recognition strengthens existing ones (McDaniel and Mason 1985).

The second research question addressed the impact of translation on eye movements and was divided into two sub-questions. The first sub-question (RQ2a) asked whether a different eye movement pattern emerged between the literal and non-literal condition. Interestingly, no significant effect of translation condition on eye movements emerged. Fixation analyses showed that (a) some translations can be given the same amount of visual processing and yet yield significantly different recall scores; (b) non-literal renderings attracted more and longer looks than literal ones, as hypothesised, but the difference is only slight and non-literal items were recalled less accurately in the recognition post-test overall. This would suggest that during L2 subtitle reading, literal translations can be as salient and attract as many looks as non-literal ones. If a literal word or structure is interesting to the viewer, for example because it creates a departure from their expectations, it will receive a number of fixations comparable to its non-literal counterpart. This finding provides evidence that different learners do find different elements of language striking. As Seilhamer (2010: 23) puts it: "The lexical or grammatical constructions that strike one learner as particularly useful or meaningful […] may not seem terribly relevant to another learner". Thus, the analysis suggests that there is more to the relationship between language processing and memory than meets the eye, and fixation measures aptly revealed the complexity of this relationship in multimodal bilingual environments such as that of audiovisuals.

Formal similarity may allow for smoother language encoding during moment-by-moment consumption of subtitled video, as evidenced by the lower fixation counts and durations, indicating less processing effort. It may also create clarity in L1-L2 word connections in the mental lexicon, allowing the learner to move on to the next subtitle without major disruptions and later facilitating recognition, whereas formal discrepancy may create less direct connections between L1 and L2 forms and the underlying concept in the mental lexicon, creating interference at the level of exact form recognition. Sometimes, however, non-literal items can be striking and be noticed by the learner, in which case they tend to be both correctly identified in the post-test and spontaneously reported in the questionnaire, leaving a richer memory trace in the mind of the learner. Thus, these quantitative results cannot be taken to mean that subtitles should be translated literally wherever possible when intended for classroom use, as it could be extrapolated from Pavesi and Perego (2008: 223-225), because salience potentially ensuing from discrepancy clearly also has a role in FLL, at least in advanced learners.

Mean fixation duration on reversed subtitles was 217 ms, a value consistent with the only two previous studies which collected this information. d'Ydewalle and De Bruycker (2007) registered mean values of 185 ms and 201 ms for reversed one- and two-liners respectively in their adult population (242 ms overall average), while in Bisson *et al.* (2014) mean fixation duration for reversed subtitles was 243 ms. The slightly lower

duration recorded herein is ascribed to the high proficiency of the participants. The IFL test assessed B2 level, but overall good performance suggests learner proficiency in this cohort might have been slightly higher, allowing them to process L2 input even more quickly and efficiently.

The second sub-question (RQ2b) asked whether there is a relationship between visual attention and recall rates. T-tests revealed significant differences in visual attention between recall accuracy and inaccuracy. As Figures 6 and 7 show, incorrectly recalled subtitles are consistently looked at less, both in terms of fixation count and duration. This may seem self-evident, but it incidentally highlights a crucial difference between reversed subtitles and bimodal input. In the former, the subtitle is the only place where the L2 appears, so failure to read an L2 string is more directly linked to failure to correctly identify its wording. In the latter, on the other hand, since L2 input appears both aurally and in writing, failure to read an L2 string could still result in its accurate identification, if its aural counterpart was processed and understood by the learner.

Finally, overall recognition scores for all participants across translation conditions were unexpectedly high (72.3%), indicating that, in most cases, subjects retained a great deal of L2 information from the subtitles, regardless of how close or distant they were from their L1 audio counterparts. Therefore, exploiting reversed subtitles in FLL has mnemonic potential, and not only at beginner and intermediate levels, as already suggested in Danan (2004), Díaz-Cintas and Fernández Cruz (2008) and Ghia (2012a), but also at more advanced stages of learning.

9. Limitations and future research

Limitations of the current study include the wide variety of word types and syntactic structures considered, the use of total fixation duration and counts only, and the collection of eye data at subtitle rather than word level. Despite these limitations, however, this paper did show that reversed subtitling has some acquisitional potential when used with true L2 learners. The importance of addressing acquisitional and behavioural measures together (Kruger 2016) also became apparent, and it is therefore hoped that experimental studies continue to use this double paradigm to investigate L2 learning stemming from subtitle use. Research will also benefit from analysing different fixation types (e.g. first- versus second-pass), saccadic movements (e.g. regressions) and other measures such as latencies or skipping rates. Finally, future studies should also include delayed post-tests, in order to assess cumulative effects and ascertain whether immediate results hold over time.

10. Conclusions

As Díaz-Cintas and Fernández Cruz note, although more empirical research is needed, there is widespread agreement that different combinations of

script and audio on the moving images can activate prior knowledge and allow participants to "practice, expand and maintain that linguistic knowledge" (2008: 205). This paper confirms that in one such combination, reversed subtitling, the script – that is to say, the text in the subtitles – is indeed processed, a great deal of L2 input is retained, and translation-specific features are noticed and can impact recognition memory. The paper thus also highlights how translation, in view of its psycholinguistic benefits, deserves a more prominent place in cognitive studies where bilingual processing and memory are concerned.

References

- **Balota, David A., Alexander Pollatsek and Keith Rayner** (1985). "The interaction of contextual constraints and parafoveal visual information in reading." *Cognitive Psychology* 17(3), 364-390.

- **Baltova, Iva** (1999). *The Effect of Subtitled and Staged Video Input on the Learning and Retention of Content and Vocabulary in a Second Language*. PhD Thesis. Toronto: University of Toronto.

- **Barni, Monica, Anna Bandini, Laura Sprugnoli, S. Lucarelli, A.M. Scaglioso, B. Strambi, C. Fusi and A.M. Arruffoli** (2009). *Linee guida CILS. Certificazione di Italiano come Lingua Straniera*. Università per Stranieri di Siena. Perugia: Guerra Edizioni.

- **Bisson, Marie-Josee, Walter J. B. Van Heuven, Kathy Conklin and Richard J. Tunney** (2014). "Processing of native and foreign language subtitles in films: An eye tracking study." *Applied Psycholinguistics* 35, 399-418.

- **Caimi, Annamaria** (2005). "Subtitling in a cognitive perspective to encourage second language learning." John D. Sanderson (ed.). *Research on Translation for Subtitling in Spain and Italy*. Alicante: University of Alicante, 65-77.

- **Caimi, Annamaria** (2012). "Cognitive insights into the role of subtitling in L2 learning." Adriana Şerban, Anna Matamala and Jean Marc Lavaur (eds). *Audiovisual Translation in Close-up: Practical and Theoretical Approaches*. Bern: Peter Lang, 113-129.

- **Čepon, Slavica** (2011). "Interlingual subtitling as a mode of facilitating incidental foreign language acquisition." *English for Specific Purposes World* 33(11): 1-37.

- **Čepon, Slavica** (2013). "Effective use of the media: Video in the foreign language classroom." *Medijska Istraživanja* 19(1): 83-104.

- **Clifton, Charles Jr., Adrian Staub and Keith Rayner** (2007). "Eye movements in reading words and sentences." Roger P.G. van Gompel, Martin H. Fischer, Wayne S. Murray and Robin L. Hill (eds). *Eye Movements: A Window on Mind and Brain*. Amsterdam: Elsevier, 341-371.

- **Cop, Uschi, Denis Drieghe and Wouter Duyck** (2015). "Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel." *PLoS ONE* 10(8): 1-38.

- **Danan, Martine** (1992). "Reversed subtitling and dual coding theory: New directions for foreign language instruction." *Language Learning* 42(4): 497-527.

- **Danan, Martine** (2004). "Captioning and subtitling: Undervalued language learning strategies." *Meta* 49(1): 67-77.

- **Daneman, Meredyth and Patricia A. Carpenter** (1980). "Individual differences in working memory and reading." *Journal of Verbal Learning and Verbal Behavior* 19: 450-466.

- **De Linde, Zoe and Neil Kay** (1999). *The Semiotics of Subtitling*. Manchester: St. Jerome.

- **Díaz-Cintas, Jorge and Aline Remael** (2007). *Audiovisual Translation: Subtitling*. Manchester: St. Jerome.

- **Díaz-Cintas, Jorge and Marco Fernández Cruz** (2008). "Using subtitled video materials for foreign language instruction." Jorge Díaz-Cintas (ed.) *The Didactics of Audiovisual Translation*. Amsterdam: John Benjamins, 201-214.

- **Dijkstra, Ton, Jonathan Grainger and Walter J.B. van Heuven** (1999). "Recognition of cognates and interlingual homographs: The neglected role of phonology." *Journal of Memory and Language* 41: 496-518.

- **d'Ydewalle, Géry and Wim De Bruycker** (2007). "Eye movements of children and adults while reading television subtitles." *Psychological Research* 71(4): 458-466.

- **d'Ydewalle, Géry and Ingrid Gielen** (1992). "Attention allocation with overlapping sound, image, and text." Keith Rayner (ed.) *Eye Movements and Visual Cognition: Scene Perception and Reading*. New York: Springer-Verlag, 414-427.

- **d'Ydewalle, Géry and Ubolwanna Pavakanun** (1996). "Le sous-titrage à la télévision facilite-t-il l'acquisition des langues ?" Yves Gambier (ed.) *Les Transferts linguistiques dans les médias audiovisuels*. Lille: Presses Universitaires du Septentrion, 217-223.

- **d'Ydewalle, Géry and Ubolwanna Pavakanun** (1997). "Could enjoying a movie lead to language acquisition?" Peter Winterhoff-Spurk and Tom H.A. van der Voort (eds). *New Horizons in Media Psychology*. Opladen: Westdeutscher Verlag GmbH, 145-155.

- **d'Ydewalle, Géry and Marijke Van de Poel** (1999). "Incidental foreign-language acquisition by children watching subtitled television programs." *Journal of Psycholinguistic Research* 28(3): 227-244.

- **Francis, Wendy S.** (2005). "Bilingual semantic and conceptual representation." Judith F. Kroll and Annette M.B. de Groot (eds). *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: Oxford University Press, 251-267.

- **Gerber-Morón, Olivia and Agnieszka Szarkowska** (2018). "Line breaks in subtitling: an eye tracking study on viewer preferences." *Journal of Eye Movement Research* 11(3): 1-22.

- **Ghia, Elisa** (2012a). *Subtitling Matters: New Perspectives on Subtitling and Foreign Language Learning*. Oxford: Peter Lang.

- **Ghia, Elisa** (2012b). "The impact of translation strategies on subtitle reading." Elisa Perego (ed.) *Eye Tracking in Audiovisual Translation*. Rome: Aracne, 157-182.

- **Godfroid, Aline, Alex Housen and Frank Boers** (2010). "A procedure for testing the noticing hypothesis in the context of vocabulary acquisition." Martin Pütz and Laura Sicola (eds). *Inside the Learner's Mind: Cognitive Processing and Second Language Acquisition*. Amsterdam: John Benjamins, 169-197.

- **Gottlieb, Henrik** (1992). "Subtitling: A new university discipline." Cay Dollerup and Anne Loddegaard (eds). *Teaching Translation and Interpreting. Training, Talent and Experience.* Amsterdam: John Benjamins, 161-170.

- **Guichon, Nicolas and Sinead McLornan** (2008). "The effects of multimodality on L2 learners: Implications for CALL resource design." *System* 36(1): 85-93.

- **Guillory, Helen Gant** (1998). "The effects of keyword captions to authentic French video on learner comprehension." *Calico Journal* 15(1/3): 89-108.

- **Hall, Christopher J.** (2002). "The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development." *International Review of Applied Linguistics in Language Teaching* 40: 69-87.

- **Hamada, Megumi and Keiko Koda** (2011). "Similarity and difference in learning L2 word-form." *System* 39: 500-509.

- **Harrington, Michael and Mark Sawyer** (1992). "L2 working memory capacity and L2 reading skill." *Studies in Second Language Acquisition* 14(1): 25-38.

- **Hervey, Sándor G.J, Ian Higgins, Stella Cragie and Patrizia Gambarotta** (2000). *Thinking Italian Translation. A Course in Translation Method: Italian to English*. London: Routledge.

- **Holobow, Naomi E., Wallace E. Lambert and Liliane Sayegh** (1984). "Pairing script and dialogue combinations that show promise for second or foreign language learning." *Language Learning* 34(4): 59-76.

- **Hyönä, Jukka and Alexander Pollatsek** (1998). "Reading Finnish compound words: Eye fixations are affected by component morphemes." *Journal of Experimental Psychology: Human Perception and Performance* 24(6): 1612-1627.

- **Inhoff, Albrecht Werner and Keith Rayner** (1986). "Parafoveal word processing during eye fixations in reading: Effects of word frequency." *Perception & Psychophysics* 40(6): 431–439.

- **Juhasz, Barbara J. and Keith Rayner** (2003). "Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29(6): 1312-1318.

- **Juhasz, Barbara J. and Keith Rayner** (2006). "The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations." *Visual Cognition* 13(7-8): 846-863.

- **Just, Marcel A. and Patricia A. Carpenter** (1980). "A theory of reading: From eye fixations to comprehension." *Psychological Review* 87: 329-354.

- **Karamitroglou, Fotios** (1998). "A proposed set of subtitling standards in Europe." *Translation Journal* 2(2). https://translationjournal.net/journal/04stndrd.htm (consulted 27.11.2018).

- **Keating, Gregory D.** (2014). "Eye tracking with text." Jill Jegerski and Bill VanPatten (eds). *Research Methods in Second Language Psycholinguistics*. New York: Routledge, 69-92.

- **Kliegl, Reinhold, Ellen Grabner, Martin Rolfs and Ralf Engbert** (2004). "Length, frequency, and predictability effects of words on eye movements in reading." *European Journal of Cognitive Psychology* 16(1-2): 262-284.

- **Koda, Keiko** (1996). "L2 word recognition research: A critical review." *The Modern Language Journal* 80(4): 450-460.

- **Koda, Keiko** (1997). "Orthographic knowledge in L2 lexical processing: A cross-linguistic perspective." James Coady and Thomas Huckin (eds). *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, 35-52.

- **Koda, Keiko** (1999). "Development of L2 intraword orthographic sensitivity and decoding skills." *The Modern Language Journal* 83: 51-64.

- **Kruger, Jan-Louis** (2016). "Psycholinguistics and audiovisual translation." *Target* 28(2): 276-287.

- **Kruger, Jan-Louis and Faans Steyn** (2014). "Subtitles and eye tracking: Reading and performance." *Reading Research Quarterly* 49(1): 105-120.

- **Kruger, Jan-Louis, Agnieszka Szarkowska and Izabela Krejtz** (2015). "Subtitles on the moving image: An overview of eye tracking." Sean Redmond and Craig Batty (eds). *Eye-Tracking the Moving Image*. *Refractory: A Journal of Entertainment Media* 25. http://refractory.unimelb.edu.au/2015/02/07/kruger-szarkowska-krejtz (consulted 10.09.2018).

- **Lambert, Wallace E., I. Boehler and Nelly Sidoti** (1981). "Choosing the languages of subtitles and spoken dialogues for media presentations: Implications for second language education." *Applied Psycholinguistics* 2(2): 133-144.

- **Lambert, Wallace E. and Naomi E. Holobow** (1984). "Combinations of printed script and spoken dialogues that show promise for beginning students of a foreign language." *Canadian Journal of Behavioural Science* 16: 1-11.

- **Laufer, Batia** (1997). "The lexical plight in second language reading. Words you don't know, words you think you know, and words you can't guess." James Coady and Thomas Huckin (eds). *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, 20-34.

- **Laufer, Batia** (2000). "Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity." *Studia Linguistica* 54(2): 186-196.

- **Leonardi, Vanessa** (2010). *The Role of Pedagogical Translation in Second Language Acquisition: From Theory to Practice*. Bern: Peter Lang.

- **Marian, Viorica** (2009). "Audio-visual integration during bilingual language processing." Aneta Pavlenko (ed.) *The Bilingual Mental Lexicon: Interdisciplinary Approaches*. Bristol: Multilingual Matters, 52-78.

- **Mitterer, Holger and James M. McQueen** (2009). "Foreign subtitles help but native-language subtitles harm foreign speech perception." *PLoS ONE* 4(11): e7785.

- **Montero Perez, Maribel, Elke Peters, Geraldine Clarebout, Piet Desmet** (2014). "Effects of captioning on video comprehension and incidental vocabulary learning." *Language Learning and Technology* 18(1): 118-141.

- **Moran, Siobhan** (2008). *The Effect of Linguistic Variation on Subtitle Reception*. MA Thesis. York: University of York.

- **Moran, Siobhan** (2012). "The effect of linguistic variation on subtitle reception." In Elisa Perego (ed.) *Eye-Tracking in Audiovisual Translation*. Roma: Aracne, 183-222.

- **Morris, Robin K.** (1994). "Lexical and message-level sentence context effects on fixation times in reading." *Journal of Experimental Psychology: Learning, Memory and Cognition* 20(1): 92-103.

- **Muljani, Djojomihardjo, Keiko Koda and Danny R. Moates** (1998). "The development of word recognition in a second language." *Applied Psycholinguistics* 19: 99-113.

- **Munday, Jeremy** (2008). *Introducing Translation Studies: Theories and Applications*. Abingdon: Routledge.

- **Newmark, Peter** (1981). *Approaches to Translation*. Oxford: Pergamon Press.

- **Pavakanun, Ubolwanna and Géry d'Ydewalle** (1992). "Watching foreign television programs and language learning." Frits L. Enge, Don G. Bouwhuis, Tom Bösser and Géry d'Ydewalle (eds). *Cognitive Modelling and Interactive Environments in Language Learning*. Berlin: Springer, 193-198.

- **Pavesi, Maria** (2002). "Sottotitoli: dalla semplificazione nella traduzione all'apprendimento linguistico." *Rassegna Italiana di Linguistica Applicata* 34(1-2): 127-142.

- **Perego, Elisa, Fabio Del Missier and Marco Porta** (2010). "The cognitive effectiveness of subtitle processing." *Media Psychology* 13: 243-272.

- **Pollatsek, Alexander and Jukka Hyönä** (2005). "The role of semantic transparency in the processing of Finnish compound words." *Language and Cognitive Processes* 20: 261–290.

- **Prince, Peter** (1996). "Second language vocabulary learning: The role of context versus translations as a function of proficiency." *The Modern Language Journal* 80(4): 478-493.

- **R Development Core Team** (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (consulted 20.08.2018).

- **Ragni, Valentina** (2017). *More than Meets the Eye. A Reception Study on the Effects of Translation on Noticing and Memorisation of L2 Reversed Subtitles.* PhD Thesis. Leeds: University of Leeds.

- **Rajendran, Dhevi J., Andrew T. Duchowski, Pilar Orero, Juan Martínez and Pablo Romero-Fresco** (2013). "Effects of text chunking on subtitling: A quantitative and qualitative examination." *Perspectives* 21(1): 5-21.

- **Rayner, Keith** (1986). "Eye movements and the perceptual span in band skilled readers." *Journal of Experimental Child Psychology* 41(2): 211-236.

- **Rayner, Keith and Susan A. Duffy** (1986). "Lexical access during eye fixations in reading: Are word access codes used to integrate lexical information across interword fixations?" *Journal of Memory and Language* 28(4): 444-461.

- **Rayner, Keith and Simon P. Liversedge** (2011). "Linguistic and cognitive influences on eye movements during reading." Simon P. Liversedge, Iain Gilchrist and Stefan Everling (eds). *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press, 751–766.

- **Sereno, Sara C., Patrick J. O'Donnell and Keith Rayner** (2006). "Eye movements and lexical ambiguity resolution: Investigating the subordinate bias effect." *Journal of Experimental Psychology: Human Perception and Performance* 32(2): 335-350.

- **Talamas, Adrienne, Judith F. Kroll and Robert Dufour** (1999). "From form to meaning: stages in the acquisition of second-language vocabulary." *Bilingualism: Language and Cognition* 2(1): 45-58.

- **Tanenhaus, Michael K.** (2007). "Eye movements and spoken language processing." Roger P.G. van Gompel, Martin H. Fischer, Wayne S. Murray and Robin L. Hill (eds). *Eye Movements: A Window on Mind and Brain*. Amsterdam: Elsevier, 309-326.

- **Van Assche, Eva, Wouter Duyck, Robert Hartsuiker and Kevin Diependaele** (2009). "Does bilingualism change native-language reading? Cognate effects in a sentence context." *Psychological Science* 20(8): 923-927.

- **Van de Poel, Marijke and Géry d'Ydewalle** (2001). "Incidental foreign-language acquisition by children watching subtitled television programs." Yves Gambier and Henrik Gottlieb (eds). *(Multi)Media Translation: Concepts, Practices, and Research*. Amsterdam: John Benjamins, 259–273.

- **Van Lommel, Sven, Laenen, Annouschka and Géry d'Ydewalle** (2006). "Foreign-grammar acquisition while watching subtitled television programmes." *British Journal of Educational Psychology* 76: 243-258.

- **Zarei, Abbas Ali** (2009). "The effect of bimodal, standard, and reversed subtitling on L2 vocabulary recognition and recall." *Pazhuhesh-e Zabanha-ye Kharej* 49: 65-85.

- **Wang, Min, Keiko Koda and Charles A. Perfetti** (2003). "Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners." *Cognition* 87: 129-149.

- **Whitley, Elise and Johnathan Ball** (2002). "Statistics review 6: nonparametric methods." *Critical Care* 6(6): 509–513.

- **Williams, Rihana S. and Robin K. Morris** (2004). "Eye movements, word familiarity, and vocabulary acquisition." *European Journal of Cognitive Psychology* 16(1-2): 312-339.

Biography

Valentina Ragni, PhD, is a researcher based at the University of Leeds, where she is working on projects in the fields of Translation Studies and Audiovisual Translation. Her work explores the potential of AVT inside and outside the foreign language (FL) classroom, through the use and the creation of subtitles. Her research interests include second language acquisition, psycholinguistics, FL teaching, translation, statistical methods and experimental technologies, such as eye tracking. She graduated *cum laude* from the School of Interpreters and Translators of the University of Trieste before moving to Leeds, where she completed a MA in Screen Translation Studies and a PhD in experimental AVT with an eye tracking study on the effects of reverse subtitles (L1 audio > L2 subtitles) on processing and memory. She teaches a number of courses at MA and BA levels and works as a freelance translator and proofreader.

Email: V.Ragni@leeds.ac.uk

Notes

[1] The example of letterbox presented to the reader comes from the film *Believe* (Ambar Pandey and Ashis Singh, 2014), which the directors kindly provided us with permission to reproduce in this form. The trailer is available here: https://www.youtube.com/watch?v=nJGiBiBwcyQ.

[2] During the RSPAN scoring procedure, it emerged that one participant had a WM much higher than everyone else who took the test, recalling all sentence-final words but two in the RSPAN test. This participant was therefore an outlier, and because the primary purpose of the RSPAN test in this experiment was to rule out possible confounds due to differences in VM, she had to be excluded from the analysis.